PB-218 716

# VALIDITY OF THE AIR QUALITY DISPLAY MODEL CALIBRATION PROCEDURE

Glenn W. Brier

Prepared for:

Environmental Protection Agency

January 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

EPA-R4-73-017

January 1973

PB 218 716

Environmental Monitoring Series

# Validity of the Air Quality Display Model Calibration Procedure

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151

Office of Research and Monitoring
National Environmental Research Center
U.S. Environmental Protection Agency
Research Triangle Park, N.C. 27711

| BIBLIOGRAPHIC DATA SHEET | 1. Report No. EPA-R4-73-017 | 2. | 3. Recipient's Accession No. PB 213 716 |
|---|---|---|---|
| 4. Title and Subtitle<br>Validity of the Air Quality Display Model Calibration Procedure | | | 5. Report Date<br>March 1973<br>6. |
| 7. Author(s)<br>Glenn W. Brier | | | 8. Performing Organization Rept. No. |
| 9. Performing Organization Name and Address<br>Glenn W. Grier, Consultant<br>Fort Collins, Colorado | | | 10. Project/Task/Work Unit No.<br>11. Contract/Grant No.<br>Special Study |
| 12. Sponsoring Organization Name and Address<br>EPA, National Environmental Research Center<br>Meteorology Laboratory<br>Research Triangle Park, North Carolina 27711 | | | 13. Type of Report & Period Covered<br>Final Report<br>14. |

15. Supplementary Notes

16. Abstracts

The study examines, from the point of view of statistical theory, the validity of the "calibration procedure" that is currently used with climatological models of multiple-source urban air pollution (such as the Air Quality Display Model), and particularly its use as a basis for predictions of air quality that would result from distributions of emissions differing from that for which the calibration was actually established. Suggestions are made that would make the procedure more meaningful.

17. Key Words and Document Analysis. 17a. Descriptors

Urban air pollution
Air quality model
Calibration
Statistical theory
Regression analysis

17b. Identifiers/Open-Ended Terms

Air pollution

17c. COSATI Field/Group

| 18. Availability Statement | 19. Security Class (This Report) | 21. No. of Pages |
|---|---|---|
| Unlimited | UNCLASSIFIED | 32 |
| | 20. Security Class (This Page) UNCLASSIFIED | 22. Price $3.00 |

EPA-R4-73-017

# Validity of the Air Quality Display Model Calibration Procedure

by

Glenn W. Brier
Consultant
Fort Collins, Colorado

Program Element No. A11009

Project Officer: Kenneth L. Calder

Meteorology Laboratory
National Environmental Research Center
Research Triangle Park, N.C. 27711

Prepared for

OFFICE OF RESEARCH AND MONITORING
NATIONAL ENVIRONMENTAL RESEARCH CENTER
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, N.C. 27711

January 1973

This report has been reviewed by the Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Contents


| | | Page |
|---|---|---|
| 1. | Introduction | 1 |
| 2. | Background | 1 |
| 3. | Statistical considerations | 3 |
| | a. Validity | 3 |
| | Statistical validity | 4 |
| | Validity and usefulness | 5 |
| | b. Regression analysis | 6 |
| | c. Assumptions in regression analysis and associated significance tests | 11 |
| 4. | Application | 12 |
| | AQDM: St. Louis | 14 |
| | GEOMET: St. Louis and Chicago | 17 |
| 5. | Suggestions for possible improvement | 22 |
| 6. | Summary and conclusions | 25 |
| 7. | References | 28 |

# Validity of the Air Quality Display Model Calibration Procedure

Glenn W. Brier

1. Introduction. This study examines, from the point of view of statistical theory, the validity of the "calibration procedure" that is currently in use by EPA for use with climatological models for multiple-source urban air pollution (such as the Air Quality Display Model (AQDM)). The primary source materials and reports used here are reports by Calder (1971), GEOMET (1972) and Woodbury (1971). Occasionally references are made to other studies in the general area of urban diffusion or air pollution, in order to illustrate some point, but no attempt is made to conduct a critical survey of the literature on the subject. In regard to statistical theory and method, two well-known texts, Fisher (1935) and Snedecor & Cochran (1969), can be used by the reader who wishes to pursue the statistical aspects in more detail.

2. Background. The climatological models, such as the AQDM described by Calder (1971), calculate the long-term average (e.g. seasonal or annual) concentration of air pollutant at an arbitrarily located receptor location by simple integration of the effects produced by the multiple-source distribution of emissions, coupled with an integration over the joint frequency of occurrence of meteorological conditions (wind direction, wind speed and atmospheric stability) that

Reproduced from best available copy.

influence the atmospheric transport and diffusion of pollutant. This meteorological frequency distribution is assumed to characterize the entire urban area for the period being considered. The "calibration procedure" then consists of comparing, for a number of arbitrarily located receptor locations, the calculated concentration values with actually observed values. From a plot of the observed values against corresponding calculated values a linear regression of observed on calculated values is then computed by standard methods. Although the slope of this straight line is frequently very different from unity, i.e., the observed and calculated values are widely different, this regression method is now standardly used as a basis for adjusting all calculated values of air quality as directly obtained by use of multiple-source urban pollution models. For a specific urban area this procedure is normally used in making studies of pollution control strategies that involve air quality comparisons for various hypothetical emissions distributions very different from that for which the calibration was established. The procedure is also used in long-range planning studies where concern is with the pattern of air quality with pollutant emissions such as might occur a decade or so in the future. In such applications the meteorological conditions (as expressed through their joint frequency function) that are an input for the model calculations are held constant and the spatial distribution of emissions is varied. The study will examine the validity of this statistical calibration procedure when used to make predictions of air quality for distributions of emissions differing from that under which the calibration was actually established. It will attempt to identify specific areas of weakness and make suggestions for further studies and improvements.

3. Statistical considerations. The approach here will be to first consider some of the general scientific problems connected with the validation and calibration of a physical model and to examine the role of statistics in this process. Specific application will then be made to the problem of evaluating the AQDM, giving attention to the questions of sampling and statistical inference, particularly as they relate to the calibration procedure and its use. Since the calibration concept is quite general, the discussion also includes a reference to the results of the GEOMET (1972) analysis of the Gaussian plume multiple-source urban diffusion model. However, the GEOMET formulation of the model considered short-period variations of emissions while the essential feature of the AQDM or Calder (1971) is that these models utilize the pollutant emission rate averaged over a long time-period.

a. Validity. The agreement of a model with observations is usually referred to as validity. If the agreement is good, the model (or theory) is considered to be true, although it is generally recognized that the word true may be misleading since any model is at best an approximate description of reality. A graphical representation by means of a scatter diagram can be made showing the relationship between the model predictions X and observations Y, as shown in Figure 1.

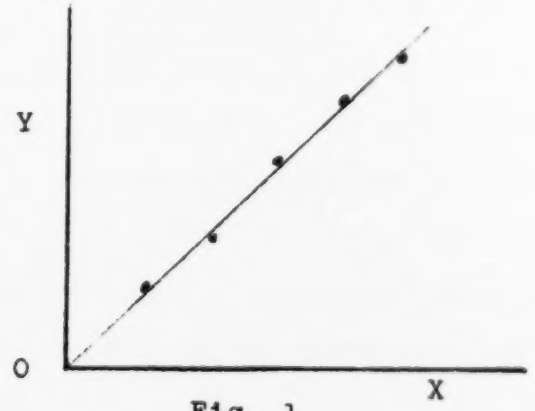


Fig. 1

Thus, for example, X may represent predictions of the motion of a planet during a specified time, using the laws of Kepler and Newton as a model. The observations made from astronomical measurement are plotted as the Y coordinate. If both X and Y are essentially free from error, the points will lie along the line Y = X (or very close to it) and the agreement will be so obvious that one might say "there is no need for statistics." The statistical problem arises because of the ever present variation in natural phenomena. The observation Y may contain a measurement error $\varepsilon$ and the model predictions X may contain an error $e$. (For the time being we will consider these errors as normally distributed with mean zero and variances $\sigma_\varepsilon^2$ and $\sigma_e^2$ respectively.) The error $e$ may result from the failure of the model to represent all the physical processes. For example, in our illustration, the model predictions might not include perturbation effects of other planets or relativity effects. Contributions to $\sigma_e^2$ will also come from errors in observing the input variances, such as the mass of the planet. If either $\sigma_\varepsilon^2$ or $\sigma_e^2$ is appreciable, the line Y = X will no longer give a good fit. However, there will be another line

$$\hat{Y} = A + BX, \tag{1}$$

(called the regression line) which will give a better fit. This line is determined by the method of <u>least squares</u> (Snedecor-Cochran 1969, p. 147), which minimizes the sum of squares of the deviations from the regression line. This is the line used for the AQDM calibration procedure.

<u>Statistical validity</u> is concerned with the interpretation of this line and the inferences that can be made correctly. The

coefficients A and B have been estimated from a sample -- a small collection of data from a much larger aggregate or population. We wish to make quantitative statements about the larger population. Now, as expressed by Fisher (1935, p. 4) "any inference from the particular to the general must be attended with some degree of uncertainty ... [but] ... the nature and degree of this uncertainty may itself be capable of rigorous expression." To be able to do this, certain conditions must be met. One requirement is stated by Snedecor and Cochran (1969) as follows " ... statements apply to the population that was actually sampled. Claims that such inferences apply to some more extensive population must rest on the judgment of the investigator or on additional extraneous information that he possesses." For example, our hypothetical astronomer examining the data of Figure 1 would feel quite justified in using the model to predict the motions of other planets. If there had been considerable departures from the line Y = X, he would not be so confident.

Another requirement for making sound inferences has to do with the appropriateness of the statistical model used, i.e., what are the assumptions and were they obeyed in this particular application. These assumptions will be discussed in Section 3c and their relevance to the AQDM calibration procedure in Section 4.

Validity and usefulness are confused sometimes and a few words might be in order. Perfect agreement between prediction and observation (Y = X) would imply validity as well as usefulness, or at least potential usefulness. On the other hand, lack of complete agreement (e.g. a correlation of 0.70 between X and Y) does not preclude the possibility of usefulness, even though it might raise questions of validity.

Woodbury (1971) says "It is clear from the moderate but not high correlation of computed and observed $SO_2$ concentration with more than half the variance accounted for, that the APCO Air Quality display model is a valuable tool for its intended purpose," and goes on to suggest that "the output of the model as it stands can be improved in accuracy with relatively modest increases in computational cost by providing the proper wind roses as primary input." It seems that a reasonable point of view would be to interpret validity in a relative rather than absolute sense, and to consider usefulness in relation to alternative methods that might be available and their cost. Statistics might be helpful in providing the investigator with a measure of how much confidence he should have in the results of a particular procedure, but he should still go ahead with attempts to develop better models.

b. Regression analysis. In the regression line given by equation (1), the coefficient B (slope) is determined by

$$B = \frac{\sum xy}{\sum x^2}, \qquad (2)$$

where $x = (X - \bar{X})$, $y = (Y - \bar{Y})$,

and $\bar{X}$ and $\bar{Y}$ represent the means of X and Y respectively.

The intercept A is given by

$$A = \bar{Y} + B\bar{X}, \qquad (3)$$

and the deviation $d_{y.x}$ of an observation Y from the regression line is given by

$$d_{y.x} = Y - \hat{Y}. \qquad (4)$$

The sum of squares of deviations, $\sum d^2_{y.x}$, is the basis for an

estimate of error in fitting the line. It is easy to show (Fisher, 1938, p. 144) that the quantity $\sum d^2_{y.x}$ can be obtained by

$$\sum d^2_{y.x} = \sum y^2 - B^2 \sum x^2 \quad , \tag{5}$$

and that the mean square deviation from regression given by

$$s^2_{y.x} = \frac{\sum d^2_{y.x}}{n-2}$$

is an unbiased estimate of $\sigma^2_\varepsilon$ (introduced in Section 3a), if we consider X free from error. The reason for the divisor n-2, two less than the number of observations, is to allow for the two degrees of freedom we have lost in fitting the two adjustable constants, A and B. All of these are necessary quantities in enabling us to correctly interpret a validation or calibration procedure (subject to certain assumptions discussed below). By computing the standard error of the constant B, we can test whether it differs from unity, or from some other value. Likewise, by computing the standard error $s_A$ of the constant A we can obtain some guidance as to whether the intercept that departs from zero is physically meaningful or not. These estimated standard errors are given by

$$s_B = \left(\frac{s^2_{y.x}}{\sum x^2}\right)^{1/2} \quad , \tag{7}$$

and

$$s_A = \left[ s^2_{y.x} \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}\right) \right]^{1/2} \quad . \tag{8}$$

For the sake of enlightenment, let us now consider some hypothetical cases similar to those that we might encounter in practice. First,

let us take the case where the model predictions are essentially perfect ( $\sigma_e^2 = 0$ ) but the observations have an error ( $\sigma_\varepsilon^2 > 0$) due to measurement and sampling problems. It can be shown (Snedecor and Cochran, 1969, p. 165) that the expected value of the slope is unity, regardless of the value of $\sigma_\varepsilon^2$ . In an actual sample of data, there will be fluctuations from unity, of course, in accordance with equation (7). This situation is suggested by Fig. 2 (reproduced here from the report by GEOMET (1972))where the seasonal mean concentrations for ten St. Louis stations were predicted by applying a model that attempts to include short-term emission variations as well as meteorological variations. (More will be said later about this application.) If the slope B turns out to be close to unity and the intercept does not differ significantly from zero, the investigator might conclude that the true relation is Y = X and that no "calibration" is needed.

Another possible situation is illustrated in Figure 3, where the regression line has a slope near unity but has an appreciable intercept A.

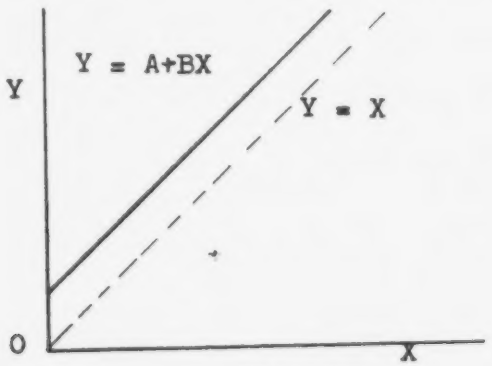


Fig. 3

A possible physical interpretation of this intercept in a diffusion model study is that A represents the background concentration. Thus, in the study by Turner et al (1971), a background concentration of

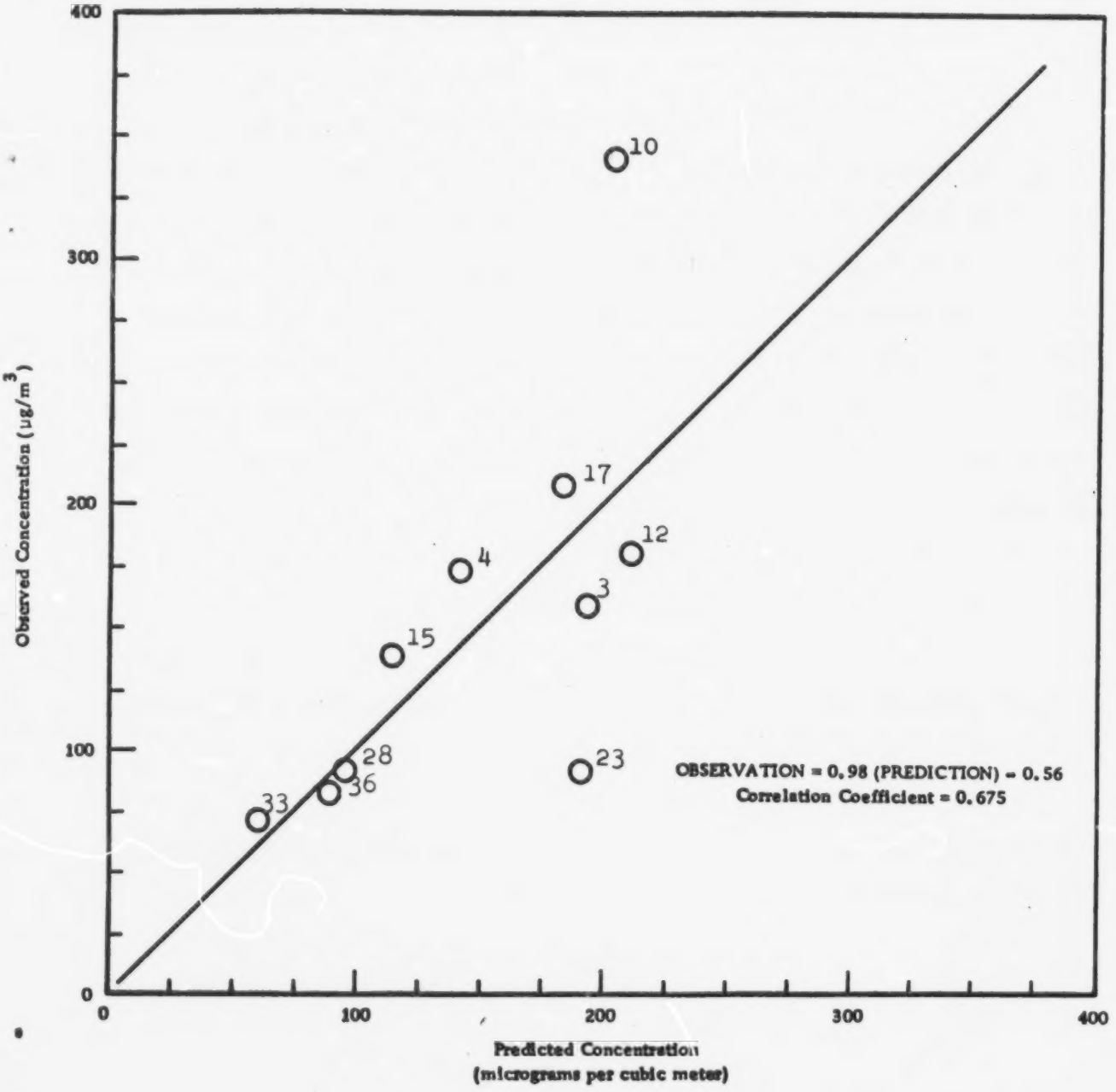


**Figure 2. Regression Analysis of Seasonal Mean Concentrations for 10 St. Louis Stations (Winter 1964-65)**

(Adapted from GEOMET, 1972)

$35 \mu g/m^{-3}$ for particulate matter was included in the dispersion models used, and this was consistent with the intercept found. A second possible explanation is that there is a systematic bias in the observations and a third possibility is that of a systematic bias in the prediction. It should be emphasized that <u>without some additional information</u>, such as measures of concentration in upwind locations, it is impossible to say which of the three explanations is the correct one. An investigator might be willing to use the regression line for calibration purposes, especially if the scatter about the line was small, but there is no assurance that background concentration or bias would be the same in different circumstances. It would be much more scientific and safer to measure the background concentration or find the sources of the bias and eliminate them.

The cases discussed above where X is free of error ( $\sigma_e^2 = 0$ ) are not very realistic ones so we now consider the more practical situation where $\sigma_e^2 > 0$ and the actual model prediction X is the sum of the <u>true</u> prediction $X^*$ plus the error $e$ . Likewise, the observation is

$$Y = Y^* + \varepsilon \quad , \tag{9}$$

where $Y^*$ is the true observation, free from error. The correlation between X and Y can be expressed in the form

$$r = \frac{1}{\left(1 + \frac{\sigma_\varepsilon^2}{\sigma_{y^*}^2}\right)^{1/2}\left(1 + \lambda\right)^{1/2}} \quad , \tag{10}$$

where

$$\lambda = \frac{\sigma_e^2}{\sigma_{x^*}^2} \quad .$$

We do not know the variances $\sigma_{x^*}^2$ and $\sigma_{y^*}^2$ of the true $X^*$ 's and $Y^*$ 's respectively, but it is clear that the correlation between X and

Y cannot be unity unless both $\sigma_\varepsilon^2$ and $\sigma_e^2$ are zero.

Also, when $\sigma_e^2 > 0$ there is an important effect on the regression coefficient B which is relevant to using a regression analysis in a calibration procedure. It can be shown (Snedecor and Cochran, 1969, p. 115) that there is a distortion downward in estimating the true structural relationship between the variables. Thus, if the true regression should be $\hat{Y} = X$, the estimated regression coefficient will not be unity, but will be given by

$$\frac{1}{1+\lambda} \quad . \tag{11}$$

Since we do not know $\sigma_e^2$, we do not know the extent of this bias. However, it will be quite small if $\sigma_e$ is relatively small, say $\sigma_e = 0.2\sigma_x$. But unless we have some way of estimating $\sigma_e^2$, we are not able to say in a particular application whether a departure of a sample regression coefficient from unity (or some other hypothetical value) is due to $\sigma_\varepsilon^2$, $\sigma_e^2$ or to some other source. In addition, the physical interpretation of the intercept as an estimate of the background concentration is questionable, since equation (3) shows that A depends upon the estimate B.

c. Assumptions in regression analysis and associated significance tests. The standard linear regression procedures and methods of statistical inference discussed above are based on a mathematical model where a number of assumptions are made. The most important of these are as follows:

(i) The regression line is linear.

(ii) The distribution of Y for a given X is Gaussian (or at least approximately).

(iii) The variance of the departures from the regression line is constant.

(iv) The $n$ sample observations must be statistically independent for valid significance tests and reliable estimates of confidence intervals.

Assumption (i) is usually made for reasons of simplicity and convenience but standard methods of curvilinear regression are available. In connection with (ii), it is not necessary that the $\varepsilon$ be normally distributed in order that $s^2_{y \cdot x}$ be an unbiased estimate of $\sigma^2_\varepsilon$. Normality is important for the standard tests of significance in regression. Other assumptions can be made instead of (iii), such as letting the variance be proportional to X. However, different estimating procedures must be used. Assumption (iv) is important if valid significance tests are to be made or confidence limits estimated. The reliability of statistical estimates depends upon the size of the sample $n$, and if the observations are not independent because of spatial or temporal correlation, then the estimated standard errors will tend to be too low. This will produce too many "significant" results and make the confidence bands appear narrower than they really should. Calibration curves have often been used without the application of significance tests and estimation of confidence levels. When they have been applied in a routine fashion, insufficient attention has been given to the validity of the statistical procedures.

4. Application. An example of an application of regression analysis to AQDM predictions is given by Calder (1971). The model was applied to the estimation of three-month average $SO_2$ concentrations

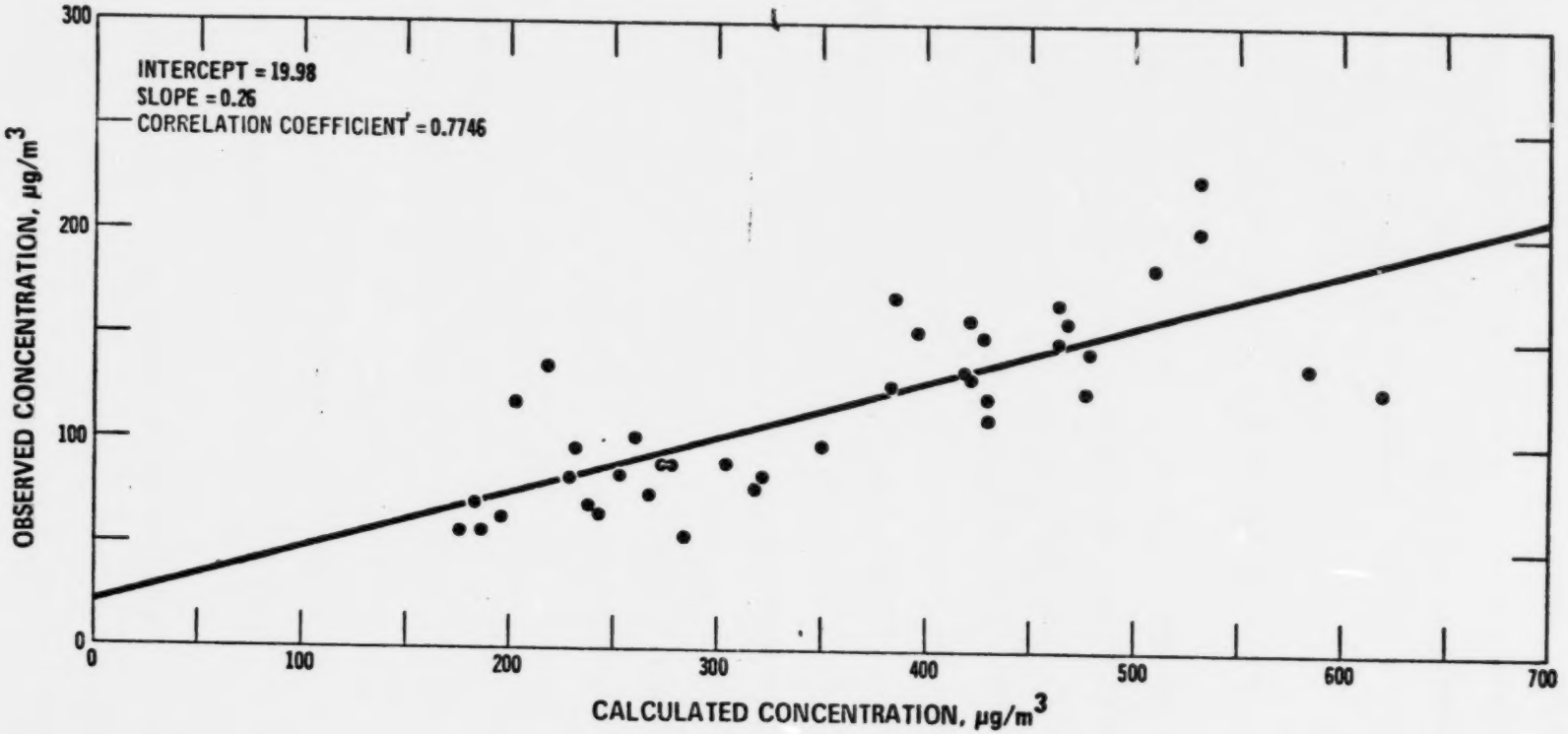


Figure 4. Regression line of observed versus calculated average $SO_2$ concentration ($\mu g/m^3$) for period December 1, 1964, through February 28, 1965. For this calculation, wind speed was assumed constant with height and $SO_2$ decay rate was assumed to be zero.

(Adapted from Calder, 1971)

in St. Louis, Missouri, during the winter of 1964-65. Figure 4, taken from that report, shows the regression of the observed on calculated values for the 40 sampling stations of the network. If Y and X denote observed and calculated values, respectively, then the regression equation is

$$\hat{Y} = 19.98 + 0.26X \qquad (12)$$

From the inspection of the graph it is obvious that the model over-calculates the concentrations and that the regression line is a much better fit to the data than the line $Y = X$, which would represent perfect agreement. The types of questions that now arise are:

(1) Can the regression line determined here be used as a calibration device for adjusting future predictions from the model when used in different seasons, locations or for a different spatial or temporal distribution of emissions?

(2) If the answer to (1) above is positive, can the regression analysis and statistics derived provide quantitative and valid measures of the confidence one can have in such extrapolations to other circumstances?

(3) If the answer to (1) above is negative, what are the conditions, if any, under which the regression line can be used for calibration purposes?

(4) What constructive suggestions can be made that might make such a calibration procedure more meaningful or effective?

The answers to such questions will depend upon whether the assumptions made in regression analysis (Section 3c) are applicable in this case and upon the more general question of validity of statistical

inferences mentioned in Section 3a. On this latter point the position of modern statistical theory and practice seems quite clear. The classic paper of "Student" (1908) opened with the following words: "Any experiment may be regarded as forming an individual or a population of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

"Now any series of experiments is only of value insofar as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong." In a public opinion poll that has as its objective the prediction of a presidential election, considerable effort is made to sample voters, and not just readers of "Time" magazine or telephone subscribers. The "Literary Digest" poll of 1936 is a tragic example of the dangers of extrapolating a sample result to a somewhat different population. This poll confidently predicted that Landon, the Republican candidate, would be elected President of the United States. Roosevelt won in a landslide, and soon afterward the publication failed. The difficulty was that a large sample was taken entirely from a few strata, completely ignoring the other strata.

In connection with the St. Louis study illustrated in Figure 4, one should feel relatively secure in using the regression line to estimate concentrations for locations in the St. Louis area where no data were available, although Kelley (1970) has suggested an alternative interpolation procedure and Woodbury (1971) has indicated that some slight improvement in agreement between observed and predicted patterns is obtained by a slight displacement and rotation of predicted pattern of isopleths. It is very questionable whether any of these procedures would be meaningful as a calibration procedure if there was a change in

the spatial or temporal distribution of emissions. Another (hypothetical) possibility would be to use the regression line for estimating concentration for another winter season for which no observed data on concentrations are available. Even assuming that emission practices have not changed, a difficulty here is that the sample from the 1964-65 winter might not be very representative of winters in St. Louis. Thus the variances and standard errors (Section 3b) are likely to be too low, since seasonal differences have not had an opportunity to contribute to the variability. In spite of this, it still might be useful if such estimates were desired. Any extension, however, to other seasons, cities or different emission patterns could not be justified on statistical grounds and the burden of proof would be on the investigator to demonstrate additional evidence or convincing reasons for making any such extension.

Let us now examine the St. Louis study in respect to the assumptions in a regression analysis. Inspection of Figure 4 suggests that there should be no problem with Assumptions (i) and (ii),[Section 3c]. There is no evidence of departure from linearity and the residuals $d_{y.x}$ do not appear to show any substantial departure from normality. Assumption (iii) might be a little bit of a problem, there being some suggestion that the residuals are larger for large values of X. In some situations there are good reasons for considering the standard deviation (or variance) of the residuals to be proportional to X, but in this particular instance it won't make much difference. Assumption (iv) is the troublesome one, for it is practically certain that the 40 observations are spatially correlated and not statistically independent. Woodbury (1971) has prepared a map of the deviations of

the observed $SO_2$ concentration from the calibrated prediction which shows non-randomness in the residuals and states that "In fact the zero isopleth can be drawn with reasonable ease and shows a relatively simple pattern." The lack of statistical independence of the sample points means that the number of degrees of freedom are considerably less than 40, and that the standard errors calculated in the usual way are lower than they should be. Keeping these points in mind, we proceed with the classical regression analysis.

The regression coefficient for the line shown in Figure 4 is $B = 0.26$ . Its standard deviation $s_B$ , computed from equation (7), is $s_B = 0.036$. This differs significantly from zero and from unity. These standard tests tell us that there is a significant relationship between X and Y, but that the line Y = X does not express it well. The intercept is $A = 19.98$ and its standard error, computed from equation (8), is $s_A = 13.25$. Although the intercept does not differ significantly from zero, it is not clear that the regression curve should be forced through zero. The standard error ($s_A = 13.25$) is rather large and as mentioned earlier, model deficiencies and background concentration would be expected to produce a positive intercept. The case for a zero intercept might be supported on physical grounds, for if there were no emissions to put into the model both the predictions and the observations would be zero. The question must remain open since there are no data (either observed or predicted) near zero to provide sufficient information.

It is also of interest to examine the results of the regression analysis of seasonal mean concentrations for the 10 St. Louis stations reported by GEOMET (1972). This study applied the Gaussian-plume type

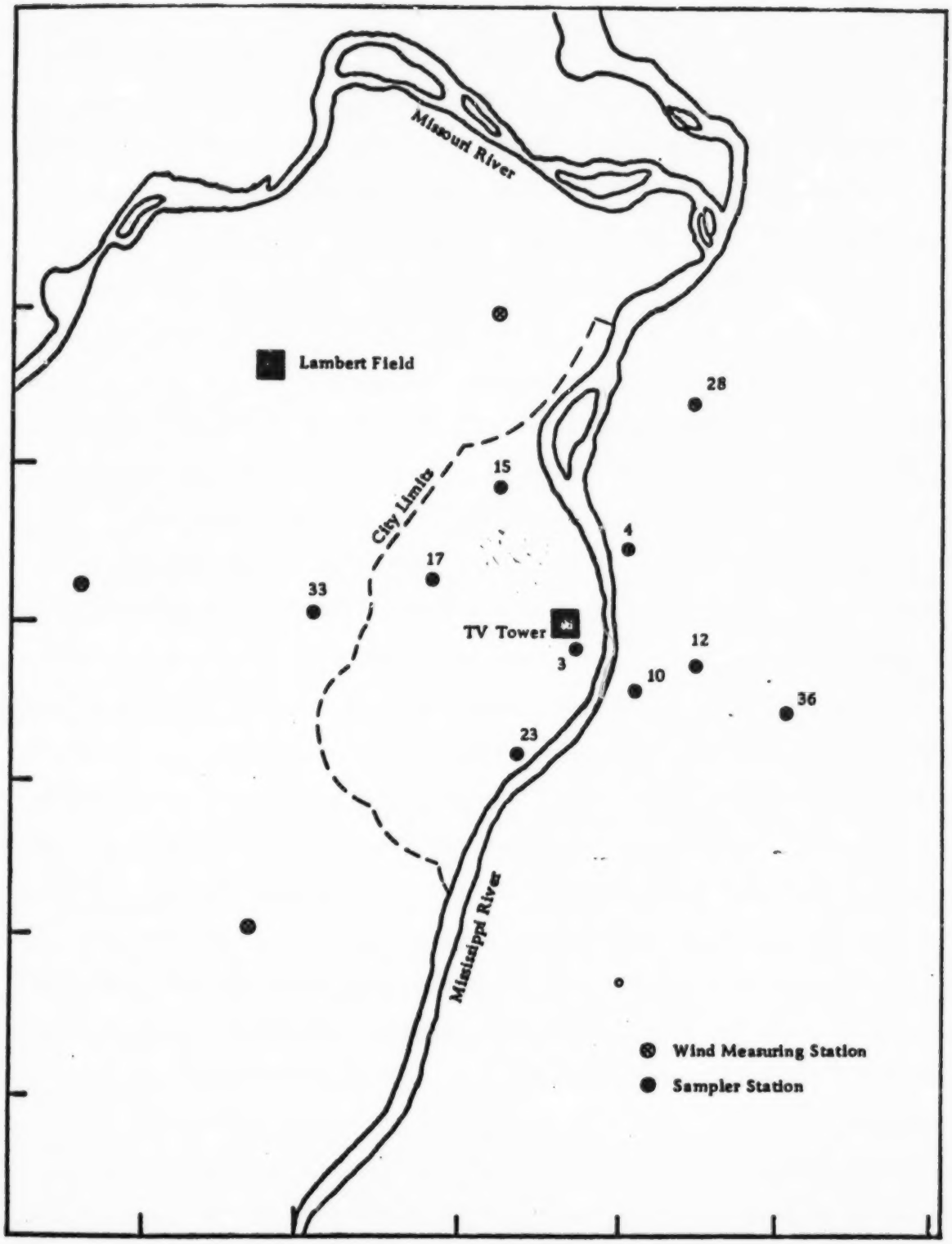


**Figure 5. Location of St. Louis Observing Stations Used in Validation Analysis**

(Adapted from GEOMET, 1972)

model to short-term emission variations. Figure 5 shows the location of the observing stations and Figure 2 shows the regression analysis. The predicted long-term concentrations appeared to show consistently good agreement with observations, as contrasted with the significant overestimation usually found in other model implementations. The regression coefficient of 0.98, with a standard error 0.38, does not depart significantly from unity and the intercept B = -0.56, with a standard error of 59.70, is within normal sampling variation of a hypothetical zero intercept. The regression coefficient near unity suggests that $\lambda$ (Section 3b) is quite small and that most of the scatter about the regression line is due to observational errors. There might be some concern about the lack of constancy of variance of the deviations about the regression line, the larger deviations being associated with large values of X. However, in this case it is not likely that it would make much difference in the slope of the regression line if a different mathematical model was used for the regression analysis. It is noted that the five highest values of X (points 3, 10, 12, 17 and 23) are all near the center of the area clustered around the TV tower.

A similar analysis was made by GEOMET (1972) for eight Chicago stations. The location of these stations is shown in Figure 6, and Figure 7 shows the regression analysis. The regression coefficient is 0.63 and its departure from unity is barely significant, using the standard tests. The standard error of the coefficient is 0.145 and this estimate is probably too low for reasons discussed in Section 3c. In fact, the greatest overprediction is point number 4, and if this point is eliminated the line Y = X (dashed line) is a good fit.

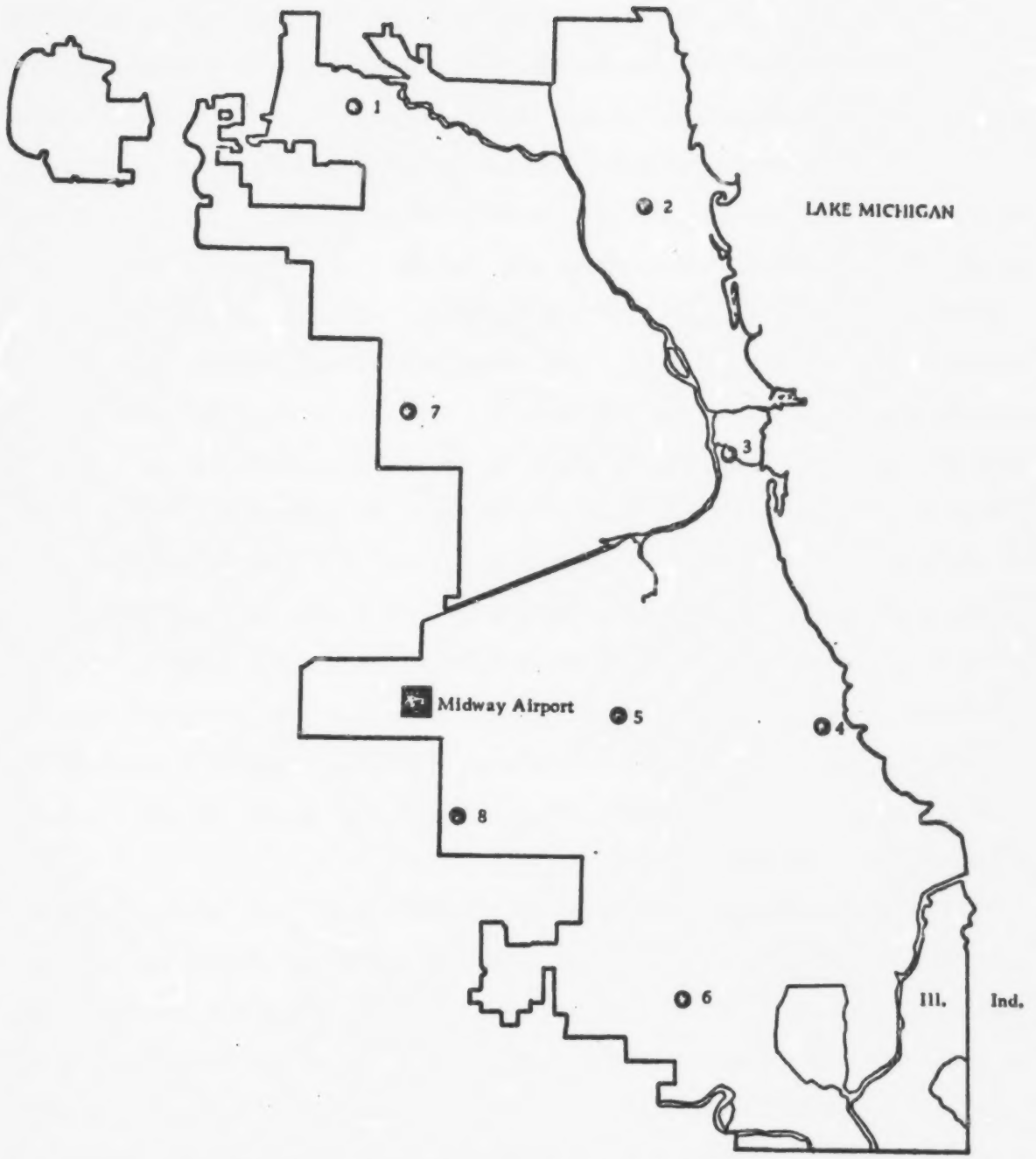


Figure 6. Location of Chicago TAM Stations Used in Validation Analysis

(Adapted from GEOMET, 1972)

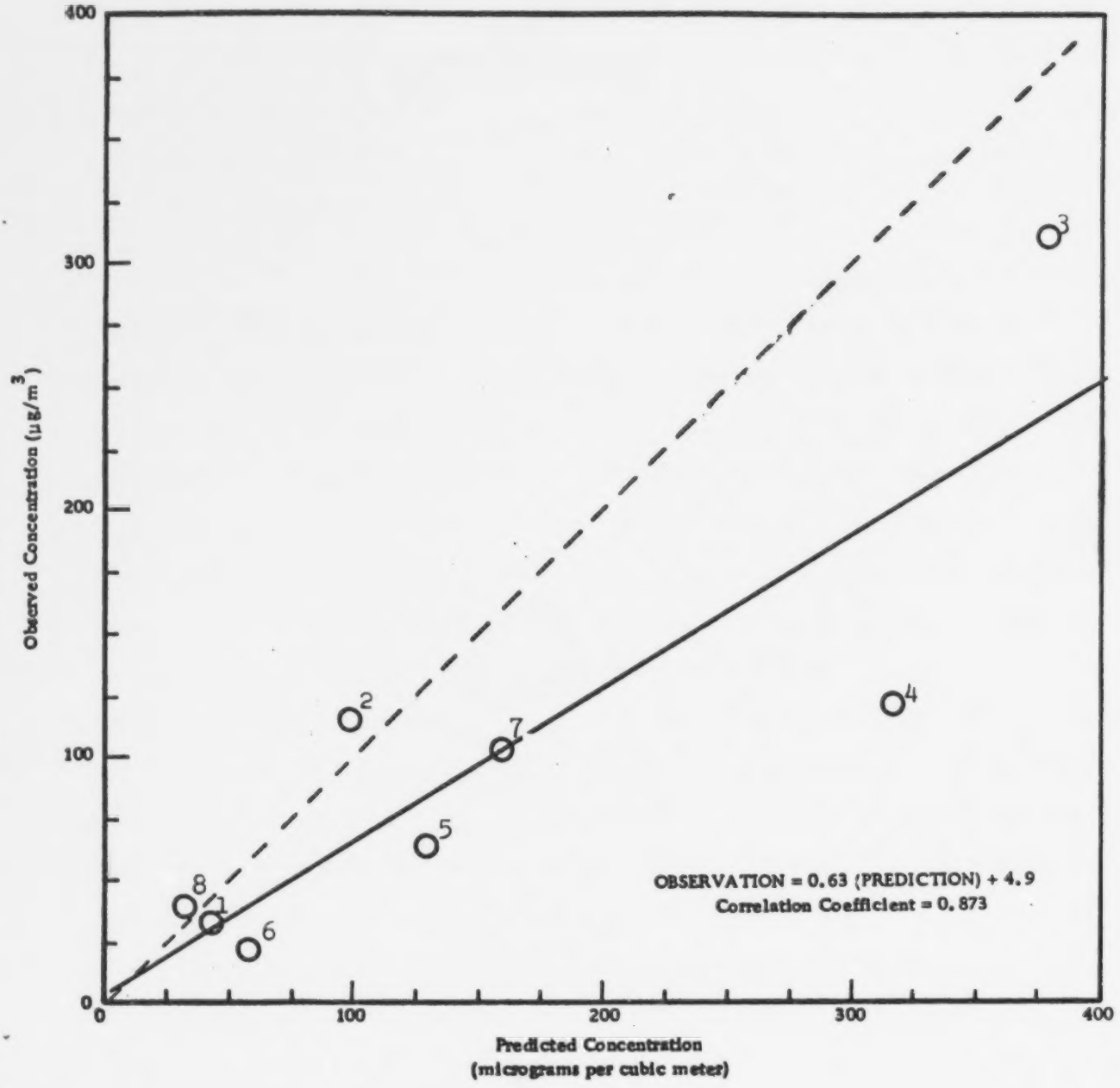


**Figure 7. Regression Analysis of Monthly Mean Concentrations for Eight Chicago Stations (January 1967)**

(Adapted from GEOMET, 1972)

Points 3 and 5 have the next largest overpredictions. These three points are geographically adjacent to one another (see Figure 6) and probably don't represent independent observations. Considering these various factors, it appears that no calibration procedure is needed and that, in fact, one might be worse off by using one in future applications.

5. Suggestions for possible improvement.

An appraisal of the results of the studies by Woodbury (1971), GEOMET (1972) and the analysis presented here indicates that a "calibration procedure" is not the answer to the problem as to how to make predictions of air quality for distributions of emissions differing from that under which the calibration of the climatological dispersion model was actually established. The valid applications of the procedure are likely to be limited to special circumstances with little practical value. However, in spite of the objections to the general application of the calibration procedure, there may be some situations where no alternative is available and where it would be desirable to consider possible changes or modifications that could make the procedure more meaningful, even though it didn't solve the bulk of the problems. One has to remember, though, that any effort put into an attempt to improve a questionable procedure might better be spent elsewhere.

One area that needs to be explored further is the matter of the air quality observations. The data represent a mixture of instrumental error, local (or transient) phenomena and the large-scale variability in pollutant. Mahoney et al (1969) recommend that at least one redundant set of sampling equipment should be employed within a sampling network, thus making it possible to get an independent estimate of the

error variance $(\sigma_\varepsilon^2)$ in the field observations. The estimate of $\sigma_\varepsilon^2$ would not only include the contribution due to instrumental error (which may be quite small) but also the sampling variations due to small scale eddies and site selection. If, in addition, an estimate of the variance $(\sigma_e^2)$ of the model errors can be obtained, equations (10) and (11) can be used to give some idea as to whether a departure of B from unity (and A from zero) is due to observational errors, model errors, or a combination. Input-output sensitivity analysis such as that reported by GEOMET (1972) provides some information on $\sigma_e^2$ but does not tell the whole story. Another point in regard to the observational data is the question of the length of the sampling period and how the average concentrations are obtained. Extremely high and temporary values of a pollutant could have undue effect on the averages since they may represent the failure for complete mixing from some local source near the receptor. It might be better to truncate the distributions or use a mathematical transformation that gives more weight to the minimum values in a time series rather than to the maximum values. This needs further study using some field observations from receptors that are located near each other.

Another direction that an improvement effort might take is to make separate calibrations for various meteorological or other stratifications such as Pasquill categories, wind directions, temperature classes, time of day, etc. Woodbury (1971) suggests making wind roses (for each Pasquill category) for each pollution source using the predicted emission as a unit for counting the wind rose frequencies. The wind roses when combined over sources could then be used as a model input separately for each observing station. Another possibility is

to make calibrations dependent upon the type of terrain or topography. Fortak (1971) reports, for example, that the same model cannot be used for Bremen as for the Frankfurt-Untermain Region where meteorological and topographical conditions differ, although there is not much difference in the emission inventories. It might be argued that if one had sufficient knowledge and data to provide calibrations for all these different categories, then it should be possible to incorporate the information into improved physical models and the need for calibration would disappear.

Improved meteorological and climatological data which are more representative of the area being studied might provide better inputs to the model. A major weakness of the model is its failure to take account of the correlation between the meteorological input parameters of the model and the actual short-term emission variations. For example, wind direction and temperature are related and affect emission. If these correlations were zero, the use of the long-term average emission rate would have greater justification. These points are discussed elsewhere in Calder (1971), Woodbury (1971) and GEOMET (1972) and need not be considered in detail here.

In the discussions of Section 3b and Section 3c the emphasis was on the mathematical model of linear regression with an intercept A and the assumption that $\sigma_\varepsilon^2$ was constant for all X. As mentioned earlier, in some situations there are good reasons for fitting a straight line through the origin, in which case the least squares estimate of the regression coefficient becomes

$$B = \frac{\Sigma XY}{\Sigma X^2} \quad , \qquad (13)$$

instead of that given in equation (2). In cases where the line goes through the origin, another common assumption is that the variance of $\varepsilon$ is proportional to X so that the regression estimate is now

$$B = \frac{\sum X}{\sum Y} = \frac{\bar{Y}}{\bar{X}} \quad . \tag{14}$$

If the standard deviation of $\varepsilon$ is proportional to X, the least squares estimate is

$$B = \frac{1}{n} \sum \left( \frac{Y}{X} \right) . \tag{15}$$

With modern high-speed computers it is easy for the investigator to calculate several of these estimates and to see whether it makes much difference when studying the relationship between X and Y. If it does make a difference, one of the estimates might make more physical sense than the others. However, in fitting a line through the origin it should be remembered that any appreciable background concentration that can be measured independently (or assumed) should be subtracted from the observed Y. Furthermore, any independent measurements or evidence regarding $\varepsilon$ can be helpful in deciding upon an appropriate regression model.

It should be emphasized that the suggestions above apply not only to the calibration problem but more generally to techniques of statistical data analysis that can be used to increase our understanding and lead to improved dispersion models.

6. Summary and conclusions. This study of the validity of the "calibration procedure" used with climatological models for multiple-source urban air pollution has considered some of the general

scientific problems connected with the validation and calibration of a physical model. Special attention has been given to the role of statistics in this process, particularly to the questions of sampling and statistical inference as they relate to the correlation and regression procedures used for calibration. A brief discussion of regression theory has been given along with some of the requirements for valid application. The results have been examined with several applications of the Gaussian-plume type models for multiple-source urban air pollution that have been reported in the literature previously. The conclusion has been reached that the "calibration procedure" is not a statistically valid procedure when used to make predictions of air quality for distribution of emissions differing from that under which the calibration was actually established. One reason for this conclusion is that coefficients defining a calibration curve have been estimated from a sample -- a small collection of data from a much larger population. From this sample, statistical theory permits inference statements only about the population that was actually sampled, and not from some other population coming from a different season, city, or a different distribution of emissions. Furthermore, these inference statements (significance tests, confidence limits, etc.) are valid only if certain requirements are satisfied. A very important assumption is that the sample data points are statistically independent. The observed data upon which calibration curves are based do not satisfy the conditions, because of spatial and/or temporal correlations. Thus it would appear that valid applications of the calibration procedure are likely to be limited to special circumstances with little practical value. Extension to other seasons, cities or different distributions

of emissions cannot be justified on statistical grounds and the burden of proof would be on the investigator to produce convincing reasons for making any such extension.

For situations where no alternative to a calibration procedure is immediately available, several suggestions have been made regarding steps that might be taken to make the procedure more meaningful. These involve such approaches as using improved measurement or sampling techniques, making separate calibrations for various meteorological or other strata, and using different statistical models. However, before pursuing these possibilities one should give serious consideration to whether the same effort spent elsewhere might not result in greater progress.

## References


Calder, K. L., 1971: "A Climatological Model for Multiple Source Urban Air Pollution", Proceedings of the second meeting of the expert panel on air pollution modelling. NATO Committee on the Challenges of Modern Society. Paris, France. July 26-27, 1971.

Fisher, R. A., 1938: "Statistical Methods for Research Workers", Oliver and Boyd, Edinburgh, seventh edition.

Fortak, H., 1971: "Mathematical Meteorological Modelling of Air Quality in the Untermain Region (Frankfurt)", Proceedings of the second meeting of the expert panel on air pollution modelling. NATO Committee on the Challenges of Modern Society. Paris, France. July 26-27, 1971.

GEOMET, 1972: "Validation and Sensitivity Analysis of the Gaussian Plume Multiple-Source Urban Diffusion Model". Final Report prepared under Contract Number CPA 70-94 for Division of Meteorology, Environmental Protection Agency, National Environmental Research Center, Research Triangle Park, N. C.

Kelley, J. H., 1970: "Calibration of NAPCA's Air Quality Display Model". Urbdata Associates, Inc., Philadelphia.

Mahoney, J. R., Maddaus, W. O., and Goodrich, J. C., 1969: "Analysis of Multiple Station Urban Air Sampling Data", Symposium on Multiple Source Urban Diffusion Models, University of North Carolina, Chapel Hill, N. C. October 27-30, 1969.

Snedecor, G. W. and Cochran, W. G., 1969: "Statistical Methods". Iowa State University Press, Ames, Iowa, sixth edition.

"Student", 1908: "The Probable Error of a Mean". Biometrika, Vol. 6, pp. 1-25.

Turner, D. B., Zimmerman, J. R. and Busse, A. D., 1972: "A Comparison of Air Pollutant Concentrations in the New York Area Calculated by two Climatological Dispersion Models".

Woodbury, M. A., 1971: (unpublished manuscript) "Evaluation of the APCO Air Quality Display Model Calibration Techniques", Duke University, Durham, N. C.